



HAL
open science

Recent Acquisition of Functional m6A RNA Demethylase Domain in Orchid Ty3/Gypsy Elements

Luis Alvarado-Marchena, Mireya Martínez-Pérez, Frederic Aparicio, Vicente Pallas, Florian Maumus

► **To cite this version:**

Luis Alvarado-Marchena, Mireya Martínez-Pérez, Frederic Aparicio, Vicente Pallas, Florian Maumus. Recent Acquisition of Functional m6A RNA Demethylase Domain in Orchid Ty3/Gypsy Elements. *Frontiers in Plant Science*, 2022, 13, 10.3389/fpls.2022.939843 . hal-03717458

HAL Id: hal-03717458

<https://hal.inrae.fr/hal-03717458>

Submitted on 14 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Recent Acquisition of Functional m6A RNA Demethylase Domain in Orchid Ty3/Gypsy Elements

Luis Alvarado-Marchena^{1†}, Mireya Martínez-Pérez^{1†}, Frederic Aparicio^{1†}, Vicente Pallas^{1**} and Florian Maumus^{2*†}

¹ Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas, Universitat Politècnica de València, Ingeniero Fausto Elio, Spain, ² INRAE, URGI, Université Paris-Saclay, Versailles, France

OPEN ACCESS

Edited by:

Jen-Tsung Chen,
National University of Kaohsiung,
Taiwan

Reviewed by:

Sergey Morozov,
Lomonosov Moscow State University,
Russia
Matej Lexa,
Masaryk University, Czechia

*Correspondence:

Vicente Pallas
vpallas@ibmcp.upv.es
Florian Maumus
florian.maumus@inrae.fr

†ORCID:

Luis Alvarado-Marchena
orcid.org/0000-0002-1867-4543
Mireya Martínez-Pérez
orcid.org/0000-0003-1864-8473
Frederic Aparicio
orcid.org/0000-0002-1586-9978
Vicente Pallas
orcid.org/0000-0003-4954-989X
Florian Maumus
orcid.org/0000-0001-7325-0527

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 18 May 2022

Accepted: 14 June 2022

Published: 04 July 2022

Citation:

Alvarado-Marchena L,
Martínez-Pérez M, Aparicio F, Pallas V
and Maumus F (2022) Recent
Acquisition of Functional m6A RNA
Demethylase Domain in Orchid
Ty3/Gypsy Elements.
Front. Plant Sci. 13:939843.
doi: 10.3389/fpls.2022.939843

Long terminal repeats (LTR) retrotransposons are transposable elements (TEs) representing major components of most plant genomes. The fixation of additional conserved protein domains in their genomes is considered a rare event in the course of their evolution. Such changes can bring novel functions and increase their fitness by playing a role in the regulation of their replicative cycle or by affecting their integration landscape so that the detection of new domains can in turn reveal important aspects of host-TE interactions. We have mined angiosperm genomes for the presence of additional domains in LTR retrotransposons. We report a lineage of large (25 kbp) Gypsy-type elements in the genomes of Phalaenopsis orchids that contain an additional open reading frame containing a 2-ODD domain with close similarity to those responsible for m6A RNA demethylase activity in AlkB proteins. By performing *in vitro* assays, we demonstrate the RNA binding capability and the demethylase activity of the Gypsy-encoded AlkB protein, suggesting it could be functional against cognate TE mRNA or any cellular RNA *in planta*. In line with recent literature, we propose that the fixation of an RNA demethylase in this lineage of LTR retrotransposons may reflect an important role for epitranscriptomic control in host surveillance against TEs.

Keywords: epitranscriptomic, LTR retrotransposons, m6A RNA methylation, Phalaenopsis (orchids), plants, transposable elements

INTRODUCTION

Transposable elements (TEs) are genetic sequences whose primary function is to mediate their own replication and integration in host genomes from all domains of life. Owing to their duplicative properties and their repetitive nature, TEs have plethora of impacts on eukaryotic genomes such as influencing their size (e.g., TEs contribute 45 and 80% of the human and maize genomes, respectively), mediating genomic rearrangements, generating genetic diversity, distributing regulatory sequences, shaping the architecture of chromosomes and chromatin, and contributing to the definition of sexual chromosomes (Kazazian, 2004). Two main classes are recognized: class I elements (also called retrotransposons) have a so-called copy-and-paste replication mechanism involving a genomic RNA (gRNA) intermediate, whereas class 2 elements have a cut-and-paste mechanism without gRNA intermediate (Wicker et al., 2007). While TE superfamilies share a conserved backbone of protein domains and key regulatory motifs in a specific organization, they have also evolved significant changes in their architecture along their evolutionary history. Variations can be observed for instance in protein domain content and striking examples of stable acquisitions of domains have been described at variable depth in the

phylogeny of TE superfamilies (Llorens et al., 2009). Such major transitions along the evolution of TE superfamilies can have profound consequences on TE replication and regulation as well as on their interaction with host organisms.

Long terminal repeats retrotransposons are Class I TEs that present long terminal repeats (LTRs) and found throughout eukaryotes. They comprise three superfamilies: Bel/Pao, Ty1/Copia, and Ty3/Gypsy, the latter two being commonly abundant in plant genomes (Wicker et al., 2007). Both superfamilies contain the same assortment of core domains typically encoded by the gag and pol genes, i.e., the GAG, protease, integrase, reverse transcriptase (RT) and ribonuclease H domains. Several examples of protein domain acquisition were reported among LTR retrotransposons. For instance, specific lineages from both Copia- and Gypsy-type elements have acquired envelope-like (ENV) genes sharing similarity to those found in retroviruses (Laten et al., 1998; Wright and Voytas, 1998). As another example, one of the major branches of the Gypsy-type superfamily, the chromoviruses, contains a CHROMO domain that was shown to interact with histone marks and to target integration toward heterochromatin (Marin and Llorens, 2000; Gorinsek et al., 2004; Gao et al., 2008). Furthermore, giant Gypsy-type LTR retrotransposons from the planarian *Schmidtea mediterranea* were found to contain a whole collection of ORFs encoding domains generally absent from this TE superfamily (Arkhipova and Yushenova, 2019). In plants, several additional ORFs were detected across different lineages of Ty3/Gypsy elements although they most often lack known conserved protein domains (Steinbauerova et al., 2011; Neumann et al., 2019; Vicent and Casacuberta, 2020).

The activity of TEs is tightly controlled by host cell machinery at different levels of their replication cycle throughout plant development. For instance, TEs are known to be transcriptionally silenced through epigenetic regulation such as DNA methylation and post-transcriptionally inactivated through different small RNA pathways (Borges and Martienssen, 2015; Matzke et al., 2015).

In different eukaryotes, such as mammals, yeasts and plants, a large variety of RNAs can be the subject of post-transcriptional modifications. N⁶-methyladenosine (m⁶A) is one of the most common and abundant modification in RNA molecules present in eukaryotes and may be referred to as the “fifth base” of RNA. Several studies have shown that m⁶A participates in different aspects of RNA biology such as the regulation of mRNA stability (Wang et al., 2014), translation (Wang et al., 2015), and protein-RNA interactions (Liu et al., 2015) although recently evidence of a direct role in promoting translation has been questioned (Zaccara and Jaffrey, 2020). In contrast to mammals, studies on the function of m⁶A modification in plants are scarcer (Arribas-Hernandez and Brodersen, 2020). Transcriptome-wide profiles in *Arabidopsis thaliana* detected the m⁶A modification in over two-thirds of the mRNAs (Wan et al., 2015). m⁶A depletion was described to increase the expression of m⁶A-targeted mRNAs involved in plant developmental control (Shen et al., 2016) whereas Anderson et al. (2018) found that m⁶A stabilizes transcripts by inhibiting ribonucleolytic cleavage directly upstream of these modification sites in *Arabidopsis*. This RNA modification has been recently demonstrated to be

important to regulate salt (Hu et al., 2021a; Zheng et al., 2021) and drought (He et al., 2021) stress tolerance, the infection cycle of a plant RNA virus (Martínez-Pérez et al., 2017), the stability of flowering-related genes (Duan et al., 2017), the expansion and ripening of tomato fruits (Hu et al., 2021b) or the yield and biomass of rice and potato plants (Yu et al., 2021). Thus, manipulation of plant RNA m⁶A methylation could be an encouraging strategy to improve plant growth and crop yield.

RNA m⁶A methylation is a reversible modification. In mammals, the protein complex known to add the methyl group to RNA (writers) includes the catalytic RNA methyltransferase METTL3, the allosteric activator METTL14, and several assisting factors such as WTAP. The demethylase enzymes (erasers) FTO and ALKBH5 have been shown to remove m⁶A from RNA (Hu et al., 2019). In plants, orthologs of METTL3 and METTL14 (MTA and MTB, respectively) and some of the assisting cofactors were characterized (Shen et al., 2016; Ruzicka et al., 2017). In fact, analysis of *Arabidopsis mta* knockdown mutants revealed that MTA is required for m⁶A mRNA methylation and essential for normal growth and development (Zhong et al., 2008; Yue et al., 2019). Furthermore, ALKBH5 orthologs found in *A. thaliana*, ALKBH9B and ALKBH10B, were shown to have RNA demethylation activity (Duan et al., 2017; Martínez-Pérez et al., 2017). Interestingly, while *alkbh9b* knockout mutants do not show differences in overall plant RNA m⁶A methylation level, its depletion results in hypermethylation of alfalfa mosaic virus (AMV) RNA and impairment of systemic infection (Martínez-Pérez et al., 2017, 2021). In addition, several studies have pointed out the varying roles of m⁶A modifications in the life cycle of animal- and plant-infecting RNA viruses (Gonzales-van Horn and Sarnow, 2017; Williams et al., 2019; Yue et al., 2022). More recently, RNA m⁶A methylation was also reported to play a role in the control of the replication of mammalian endogenous viruses and retrotransposons elements in contrasting ways. Indeed, mRNAs derived from endogenous retroviruses are methylated and the levels of m⁶A RNA methylation negatively affect their accumulation in mouse embryonic stem cells (ESCs) (Chelmicki et al., 2021). On the other hand, mRNAs of young LINE-1 retrotransposons are strongly methylated but positively promote LINE-1 retrotransposition in mouse ESCs (Xiong et al., 2021). While the control of TEs by balancing RNA m⁶A methylation levels appears as an emergent, boiling field of investigation, there is no clue at this stage how this modification may affect different types of TEs and whether it also plays a role in plants.

In this study, we explore the protein domain content of LTR retrotransposons across angiosperm genomes and found several previously unreported domains in Gypsy-type elements. Most intriguingly, we discovered a family of Gypsy-type elements in orchids that contain a functional RNA m⁶A demethylase acquired from the AlkB protein family.

MATERIALS AND METHODS

Identification of Protein Domains

We first searched for reverse transcriptase in 121 plant genome assemblies (**Supplementary Table 1**) representing about 90 Gb

in total). To this end, we first prepared a query library by clustering the 165 RT amino acid sequences of Gypsy and Copia reference elements taken from Gypsy database version 2 (Llorens et al., 2011) at a depth of 40% identity and we selected one representative sequence per cluster ($n = 35$). We then used the resulting RT library to search for homologous regions in target genomes with tBLASTn from ncbi blast + package. All overlapping hits on genomes were merged and the corresponding FASTA sequences within the expected size range (520–840 bp) were extracted ($n = 1,949,052$). To avoid sparing unnecessary computational time, during following steps, RT sequences from each species were clustered at a threshold of 95% identity using MMseqs2 (Steinegger and Soding, 2017) and a maximum of 50 sequences per group were selected for downward analysis. The genomic positions of RT coding regions were extended to 5 kb upstream and downstream and the corresponding sequences were extracted ($n = 1,456,640$). Extended hits were then clustered using mmseqs2 (with parameters `-c 0.5 -max-seq-len 15000`), and the groups containing at least 5 sequences were aligned with MAFFT (Katoh et al., 2002). A consensus sequence was then generated for each sequence alignment through the modules “msa2profile” (with parameters `-match-mode 1 -match-ratio 0.5`) and “profile2consensus,” resulting in 25,565 consensus sequences. Each consensus was compared to the CDD database (Marchler-Bauer et al., 2010) using MMseqs2. The classes of domains that are commonly found across LTR retrotransposons were filtered (belonging to Gag, protease, reverse transcriptase, ribonuclease, integrase, and Chromo categories) to obtain the names of uncommon conserved protein domains detected in at least 5 consensus established from a same species (i.e., corresponding to a TE family). To address the fraction of the consensus sequences representing LTR retrotransposons, we compared each one to a library of reference aa RT sequences from Copia, Gypsy, DIRS, endogenous retroviruses, Caulimoviridae, and LINEs using BLASTx. The consensus corresponding to LTR retrotransposons were identified from their best hit (highest bit score) against RT from Copia or Gypsy. Intrinsically disordered regions (IDRs) within Gypsy_Pa_2799 ORF2 were predicted using PrDOS with default settings (Ishida and Kinoshita, 2007).

Phylogenetic Analysis

To build the reverse transcriptase tree, we extracted the RT domain from Gypsy_Pa_2799 ORF1 and combined it with those from 96 reference elements obtained from the Gypsy database (Llorens et al., 2011) to produce a multiple sequence alignment using MAFFT (Katoh et al., 2002) with the option “`-maxiterate 1000`.” The resulting alignment was submitted to IQ-TREE (Minh et al., 2020) for model testing (best-fit model: LG + R6 chosen according to BIC) and phylogenetic analysis with 100 bootstrap replicates. For the AlkB tree, we first retrieved the sequences from eight predicted human AlkB proteins (Kurowski et al., 2003). The human proteins were used as queries to identify and collect homologs in *A. thaliana*. We then combined the 2-ODD domain from Gypsy_Pa_2799 ORF2 to these references sequences to produce a multiple sequence alignment using MAFFT with the options “`-maxiterate 1000 -dash -alignonly`.” The

resulting alignment was trimmed manually at the extremities to keep only the 2-ODD domain and then submitted to IQ-TREE for model testing (best-fit model: LG + R3 chosen according to BIC) and phylogenetic analysis with 100 bootstrap replicates.

Characterization of Gypsy-Type Elements Encoding the 2-ODD Domain

LTRharvest (Ellinghaus et al., 2008) was launched on *P. equestris* and *P. aphrodite* genomes using the options “`-maxdistltr 20000 -maxlenltr 3000 -mindistltr 5000`.” For each genome, the LTRharvest predictions were clustered using MMseqs2 (with parameter `-c 0.5`) (Steinegger and Soding, 2017) and the clusters containing at least five sequences were aligned with MAFFT (Katoh et al., 2002). A consensus sequence was then generated for each sequence alignment through the modules “msa2profile” (with parameters `-match-mode 1 -match-ratio 0.5`) and “profile2consensus.” Each consensus was compared to the CDD database (Marchler-Bauer et al., 2010) using MMseqs2 and those containing the 2-ODD domain (pfam13532) were selected ($n = 15$) and aligned using MAFFT. The *P. aphrodite* genome coverage was obtained by using these 15 consensus sequences as a library with RepeatMasker where accepted divergence between target and query was limited to 20% (`-div 20`) tRNA.

Purification of ALKB Proteins

The open reading frame of the putative Gypsy_Pa_2799 protein and the atALKBH9B gene as well as the paALKBH5-like His448Ala-mutant were chemically synthesized for subsequent periplasmic production strategy (signal peptide present in the N-terminal of the proteins) and cloned into pET28 plasmid fused to a Histidine tag at its C-terminal (ProteoGenix, Schiltigheim, France). Proteins were expressed in BL21 *Escherichia coli* cells and purified by immobilized metal affinity chromatography (IMAC) in native conditions following manufacturer recommendations.

Electrophoretic Mobility Shift Assays

Nucleic acid-protein interactions were analyzed by electrophoretic mobility shift assay (EMSA). For these assays 5 ng of DIG-labeled riboprobe of 5'UTR of AMV RNA 3 transcripts was denatured (5 min at 85°C) and cooled at room temperature for 15 min. Different amounts of purified His:paALKBH5-like fusion protein (0–360 ng) were added and incubated for 30 min at 4°C in 10 µL of binding buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl and 50% glycerol), and 1U of RiboLock RNase inhibitor (Thermo Fisher Scientific). The samples were then resolved in 1.5% TAE agarose gels, capillary transferred to positively charged nylon membranes (Hybond-N, GE Healthcare, Amersham) and fixed by UV. Finally, localization of the DIG-labeled probes was achieved by development with an anti-DIG system (Roche Diagnostics) according to the manufacturer's protocol. Northwestern assays were carried out as previously described (Pallas et al., 1999). Briefly, purified proteins (GST, His:paALKBH5-like or its mutant variant H448A) were electrophoresed in 12% SDS/PAGE and transferred to

nitrocellulose membranes. Membranes were incubated overnight at 4°C in Renaturing Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1 M NaCl, 0.05% Triton X-100, 1X Blocking Reagent, Roche). After this, membranes were incubated with 20 mL of the same buffer containing 50 ng/mL of the AMV sgRNA 4 labeled with digoxigenin for 3 h at 25°C.

In vitro Demethylation Assays

The m6A demethylase activity assay was performed by incubating 0.75 µg of m6A monomethylated ssRNA (Dharmacon, Inc., 5'-AUU GUC A[m6A]C AGC AGC-3') and 2.5 µg of protein for 3 h at 37°C in a reaction mixture containing 50 mM HEPES buffer (pH 7.0), 10 µM α-ketoglutarate, 100 µM L-ascorbic acid ascorbate and 20 µM (NH₄)₂Fe(SO₄)₂·6H₂O. Reactions were quenched by heating them at 95°C for 10 min. Digestion of nucleotides to nucleosides and the m6A/Adenosine (m6A/A) quantification by ultra-performance liquid chromatography–photodiode detector–quadrupole/time-of-flight–mass spectrometry (UPLC-PDA-TOF-MS) analysis of ssRNA oligonucleotide and total RNA were performed as described by Martínez-Pérez et al. (2017). Proteins used in this assay were the RNA-demethylase atALKBH9B (Martínez-Pérez et al., 2017) as positive control, the *P. aphrodite* Gypsy-element that codes a hsALKBH5-like (paALKBH5-like) and the paALKBH5-like His448Ala-mutant. This mutation affects the putative motif involved in Fe (II)-binding motif/iron core (H448xD...H), characteristic of the 2OG-FeII-dependent dioxygenase superfamily and the AlkB family (Lu et al., 2014). Three independent replicas were made with each of the proteins under study.

RESULTS

Screen for Non-canonical Domains

We searched for non-canonical protein domains embedded in LTR retrotransposons across 121 angiosperm genomes (Supplementary Table 1). To this end, we have first established a generic methodology to build a library of consensus sequences representing LTR retrotransposons from each plant genome. In brief, we identified genomic regions containing LTR retrotransposon sequences using the reverse transcriptase domain as a marker. The regions sharing high sequence identity were clustered and the sequences from each cluster were aligned to infer a consensus sequence. In total, we generated 25,565 consensus sequences, including 4,811 and 20,288 corresponding to Copia- and Gypsy-type elements, respectively (Supplementary Data). In addition, 423 consensus sequences were classified as endogenous Caulimoviridae and five were classified as LINE-type non-LTR retrotransposons. We have next scanned each consensus sequence for the presence of conserved protein domains by searching for profiles from the CDD database (Marchler-Bauer et al., 2010). After filtering those domains typically found in LTR retrotransposons, including Chromo domain models found in Chromoviruses, we have extracted the list of putative supplementary domains that were detected (Figure 1). Here, we focused on domains found in at least five

consensus sequences from a given species (hereafter referred to as the “n5” filter) to empirically select evolutionary stable fusion events and to avoid describing potentially less meaningful fusions (charted on Supplementary Figure 1).

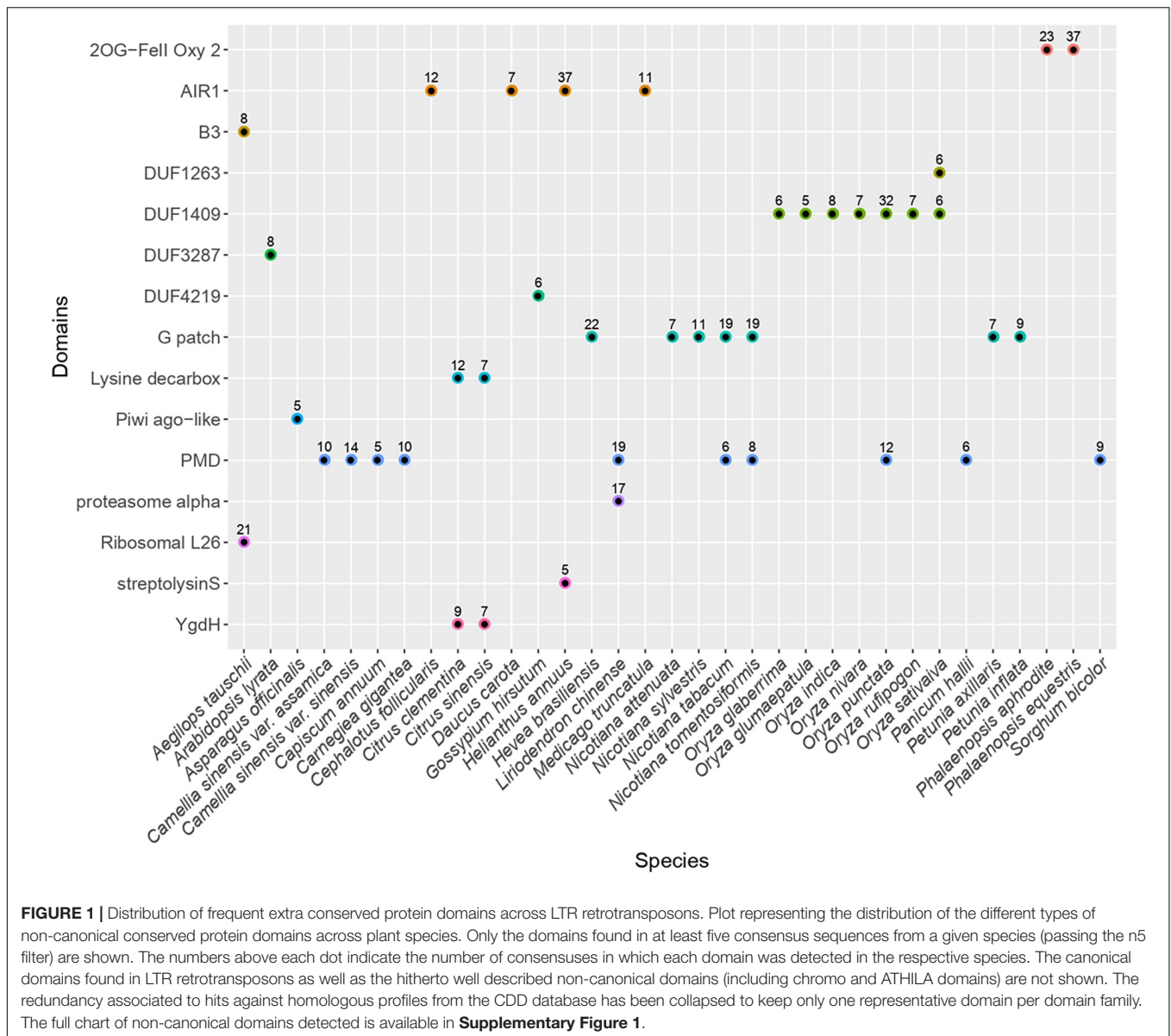
Applying this method, we detected 15 domains that are not commonly found in LTR retrotransposons, all being detected in Gypsy-type elements as opposed to Copia-type. Among the plant genomes analyzed here, seven atypical domains were found across several species, the most frequent being PMD, G-patch and AIR1.

The PMD domain (Plant Mobile Domain, PF10536), which function is unknown was the most frequent as it was detected in consensus sequences from ten plant species belonging to monocotyledons and dicotyledons (40 species when skipping the n5 filter). Though not considered a canonical Gypsy-encoded domain, PMD was previously described to associate with plant Gypsy-type elements (Steinbauerova et al., 2011) and MULE DNA transposons (Babu et al., 2006). In *Arabidopsis*, at least two PMD-containing proteins (MAIN and MAIL1) play a role in transcriptional gene regulation and TE silencing (Ikeda et al., 2017; Nicolau et al., 2020).

The G-patch domain (PF01585), containing a glycine-rich motif, is found in several eukaryotic RNA-processing proteins and is hypothesized to mediate binding to RNA. Here, this domain was found in LTR-RT consensus sequences from seven species (16 species when skipping the n5 filter) distributed across different clades of dicotyledons including the Rosids *Hevea brasiliensis* (order Malpighiales), the Asterids *Chenopodium quinoa* (order Caryophyllales), and *Nicotiana sp.* and *Petunia sp.* (order Solanales). Interestingly, the G-patch domain is also found in type D retroviral polyproteins (Aravind and Koonin, 1999).

Similarly, the AIR1 domain (Arginine methyltransferase-interacting protein), which contains a Zn-finger motif, was found in four different species (22 species when skipping the n5 filter) including the Asterids *Daucus carota* (order Apiales) and *Helianthus annuus* (order Asterales) as well as the Rosids *Medicago truncatula* (order Fabales) and *Cephalotus follicularis* (order Oxalidales). The AIR1 domain overlaps with the DNA binding universal minicircle sequence binding protein (UMSBP) domain which was previously described in some giant Gypsy-type LTR retrotransposons (Arkhipova and Yushenova, 2019) as well as in some non-LTR retrotransposons (Saint-Leandre et al., 2019).

Besides these most frequent domains, others were found in consensus sequences derived from a narrower range of species. For instance, the DUF1409 domain (PF07197), Domain of Unknown Function, was found in consensus sequences derived from seven different species from the genus *Oryza*. The taxonomic distribution of this domain appears to be restricted to the Poaceae family. Likewise, in line with a previous report (Steinbauerova et al., 2011), the YgdH domain was detected in consensus sequences obtained from two species of the *Citrus* genus, *Citrus sinensis* and *Citrus clementina* as well as in *P. trichocarpa* when n5 filter is disabled. YgdH domain is otherwise found in LONELY GUY protein family whose members play a pivotal role in regulating cytokinin activity in *Arabidopsis* (Kuroha et al., 2009) and the cell cycle in the green alga *Chlorella variabilis* (Nayar, 2020).



In addition, Ribosomal protein L26 domain was found in consensus sequences from *Aegilops tauschii* only. It contains the KOW RNA-binding motif which is shared so far among several families of ribosomal proteins (Kyrpides et al., 1996). Still from the sole *A. tauschii* genome but in different consensus, we detected the B3 domain, a DNA binding domain found in plant transcription factors with diverse functions in plant growth and development. We also identified LTR-RT consensus sequences in *Asparagus officinalis* that contain the second linker region (ArgoL2) found in PIWI- and Argonaute-like proteins. It contains residues that were described to contribute to direct the path of the phosphate backbone of nucleotides 7 through 10 in 20-nucleotides small RNAs (Faehnle et al., 2013).

Last but not least, we detected a protein domain with similarity to 2-oxoglutarate/Fe(II)-dependent dioxygenases (2-ODDs) superfamily (cl/PFAM13532) in 38 and 23 consensus

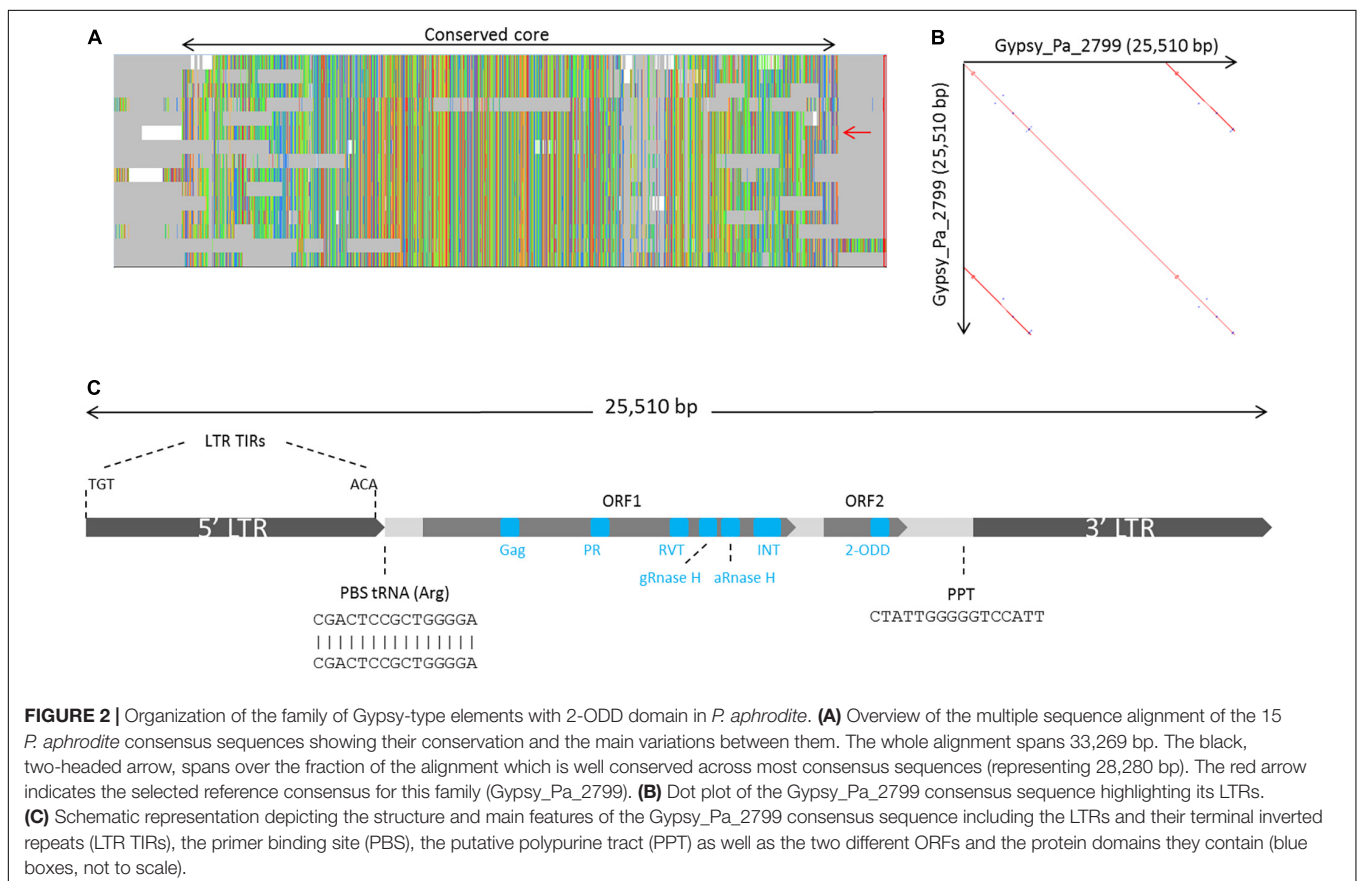
sequences established from the genome of the orchids (family Orchidaceae) *Phalaenopsis equestris* and *Phalaenopsis aphrodite*, respectively (**Supplementary Data**). Such domain was not detected in the consensus sequences derived from any other plant genome selected for this screen (even without the n5 filter), including those from several other orchids (*Dendrobium catenatum*, *Dendrobium huoshanense*, *Gastrodia elata*, *Apostasia shenzhenica*, and *Vanilla planifolia*). 2-ODDs compose a large superfamily of non-heme iron-containing domains that catalyze highly diverse oxidative reactions including hydroxylation, demethylation and desaturations in bacteria, fungi, plants, and metazoa (Herr and Hausinger, 2018). In plants, 2-ODDs are involved in a wide range of biological processes, including DNA repair, RNA and histone demethylation, hormone biosynthesis and the production of secondary metabolites (Farrow and Facchini, 2014; Kawai et al., 2014; Martínez-Pérez et al., 2017).

Phalaenopsis 2-ODD-Containing Elements Are Among Giants

We further investigated the structure and classification of the Phalaenopsis 2-ODD-containing elements. From our initial screen for LTR-RT elements across plant genomes, none of the 2-ODD-containing consensus sequence reconstructed from Phalaenopsis genomes presented long terminal repeats, indicating that those sequences might be truncated. This artifact is methodologically possible in case of large LTR-RT elements as our approach limits the reconstruction of consensus sequence to the regions 5 kb upstream and downstream of RT domains. We used LTRharvest software (Ellinghaus et al., 2008) to investigate the structure of full length corresponding elements. LTRharvest predictions were clustered on a similarity basis and the sequences from each cluster were aligned to derive consensus sequences. Following this approach, we obtained 254 and 178 consensus sequences from the *P. aphrodite* and *P. equestris* genome assemblies, respectively. After searching for conserved protein domains across these sequences, we identified 15 consensus sequences encoding the 2-ODD domain in *P. aphrodite* (Supplementary Data) but none in *P. equestris*, probably due to lesser contiguity level in this genome. The alignment of the 15 *P. aphrodite* consensus sequences revealed a core of high sequence conservation with some consensus sequences presenting structural variations (Figure 2A). We selected one sequence which covers the whole core alignment as representative (hereafter referred to as Gypsy_Pa_2799) to

analyze its structural features. The Gypsy_Pa_2799 consensus sequence is 25,510 base pairs (bp) long and comprises 6,425 bp LTRs at the extremities (Figure 2B). The internal sequence encodes two open reading frames (ORFs): ORF1 (2640 aa) contains all the conserved protein domains typically found in Gypsy-type elements in a single polypeptide while ORF2 (565 aa) carries the 2-ODD domain at its C-terminus (Figure 2C). We then addressed the overall contribution of the 2-ODD containing elements to the *P. aphrodite* genome assembly. Using the fifteen selected consensus sequences derived from LTRharvest as input library (Figure 2A) for RepeatMasker annotation, we obtained approximately 88 Mb coverage, i.e., 8.6% of the genome assembly size (9.7% when excluding assembly gaps).

We next investigated the phylogenetic relationship of 2-ODD-containing elements with other Gypsy elements. After extracting the aa RT domains from Gypsy_Pa_2799, we found that they are highly conserved, showing above 90% identity with each other, so we selected a single RT sequence as representative. We aligned this sequence with RT sequences taken from a diversity of reference Gypsy-type elements obtained from the Gypsy database to address their evolutionary relationships. The phylogenetic tree indicates that RT domains from 2-ODD-containing elements are sister to the Ogr element from *Pisum sativum* (Neumann et al., 2003) and nested within the Tat family (Figure 3). This phylogenetic placement is in line with several features of the orchid elements. First, the relatively large size compared to other



Gypsy-type elements, largely attributed to the size of LTRs, is typical of Ogre elements. Second, the primer binding site (PBS) is similar to that observed across Ogre elements, i.e., antisense to tRNA-Arginine. In addition, as reported across elements from the Tat family (Ustyantsev et al., 2015), the orchid elements also present a dual RNase H domain. However, important structural differences with Ogre elements can be observed as the Gag and Pol polyproteins are separated by a putative intron in the latter while they form a single ORF in Gypsy_Pa_2799 and as Ogre elements typically lack the additional ORF encoding the 2-ODD domain.

Phalaenopsis 2-ODDs Are Homologous to Plant RNA Demethylases

To address its putative function, we first compared the Gypsy-encoded ORF2 containing the 2-ODD domain to the NCBI

protein database and noticed the highest similarity against plant proteins belonging to the alpha-ketoglutarate-dependent hydroxylase (AlkB) superfamily, which contains the 2-ODD fold. The TE-encoded 2-ODD domain was translated from consensus Gypsy_Pa_2799 and aligned with those from human and *Arabidopsis* reference proteins to determine its position in the AlkB superfamily using phylogenetic analysis. The resulting tree shows that the Gypsy-encoded AlkB domain is nested within *Arabidopsis* AlkBH5 homologs in a clade being sister to human AlkBH5 (**Figure 4A**). The 2-ODD-containing ORF2 from *Phalaenopsis aphrodite* Gypsy elements therefore represents a hypothetical AlkBH5 homolog and it is named paALKBH5-like hereafter. It is interesting to denote that, similar to hsALKBH5, some of the most closely related orthologs in *Arabidopsis*, atALKBH9B (AT2G17970) and atALKBH10B (AT4G02940), have been shown to bind RNA and to have m6A demethylase activity (Hu et al., 2019). We next addressed whether the

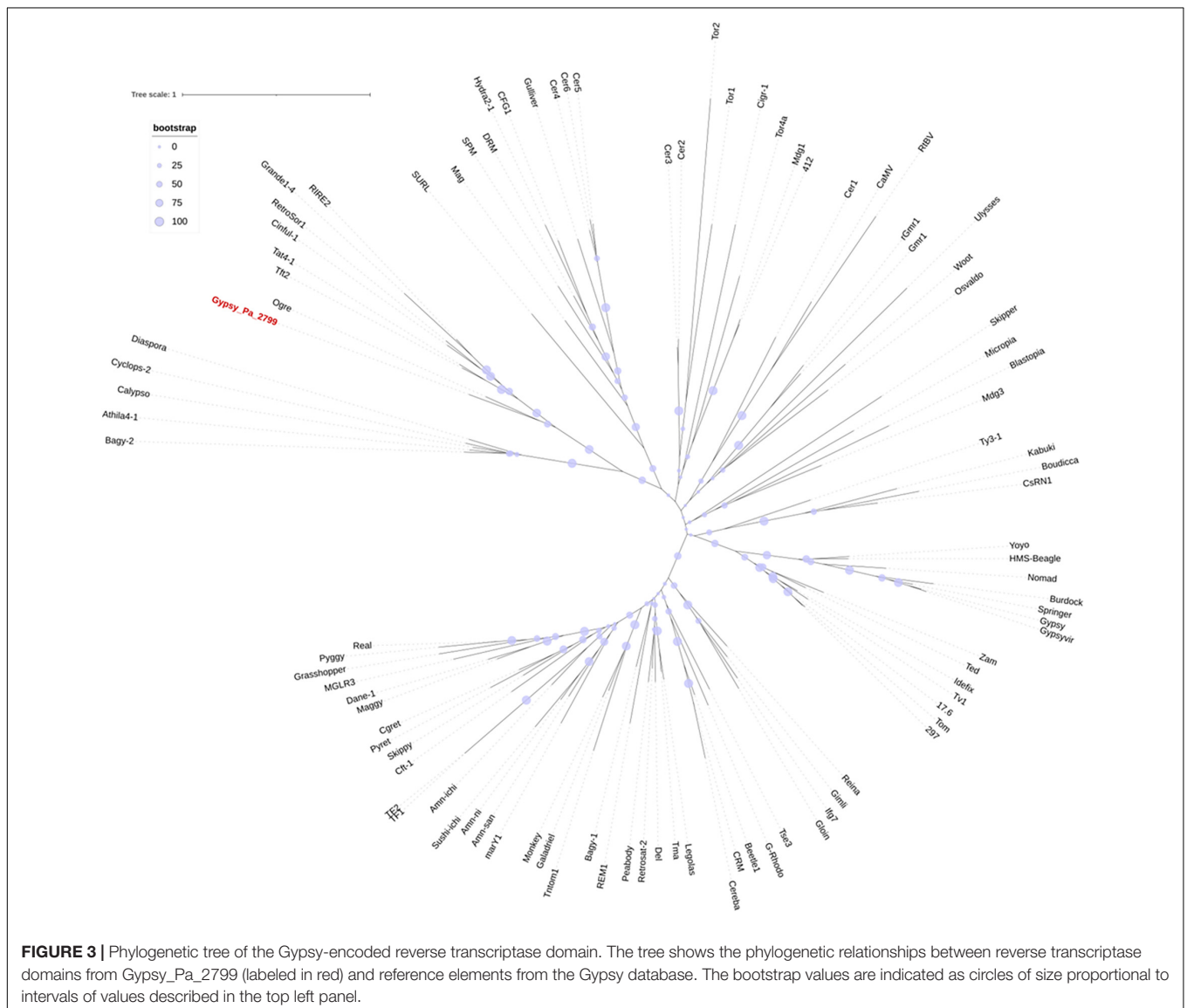


FIGURE 3 | Phylogenetic tree of the Gypsy-encoded reverse transcriptase domain. The tree shows the phylogenetic relationships between reverse transcriptase domains from Gypsy_Pa_2799 (labeled in red) and reference elements from the Gypsy database. The bootstrap values are indicated as circles of size proportional to intervals of values described in the top left panel.

sequences of TE-encoded proteins might conserve the amino acid residues necessary for this activity. The crystal structure of human AlkBH5 protein has revealed the residues binding the metal ion and 2-oxoglutarate (2-OG) cofactors and protein alignment indicated the strong conservation of these residues across the AlkB superfamily (Feng et al., 2014; Xu et al., 2014). Here, aligning the Gypsy-encoded 2-ODD domain with those of reference homologs revealed that all the cofactor-binding residues are conserved in the TE-encoded protein (**Figure 4B**). Thus, we observed that the known HxD/E...H and RxxxxR motifs (indicated as red and purple arrows, respectively, in **Figure 4B**) involved in the binding to iron [Fe (II)] and 2-OG (Bratlie and Drablos, 2005; Lu et al., 2014) are present in the paALKBH5-like protein. As well, the N and Y residues involved in 2-OG stabilization are also conserved (indicated as yellow arrows in **Figure 4B**). Furthermore, the motif YNF potentially involved in the binding to m6A and an arginine residue essential for hsALKBH5 activity (Feng et al., 2014) are also conserved (residues are highlighted in fuchsia and green, respectively, on **Supplementary Figure 2**).

Gypsy-Encoded 2-ODD Domain Has m6A Demethylase Activity Against ssRNA

To assay the m6A demethylation activity of the *P. aphroditae* Gypsy-encoded 2-ODD domain, the paALKBH5-like protein was amplified and cloned in a bacterial expression plasmid fused to a Histidine tag at its C-terminal part. As positive control, *Arabidopsis* ALKBH9B protein was cloned and expressed in the same conditions. This protein was previously found to present m6A demethylation activity toward ssRNA (Martínez-Pérez et al., 2017). Thus, m6A monomethylated ssRNA were incubated with or without the recombinant proteins. Afterward, ssRNA was digested to single nucleoside and the resultant products were analyzed by ultra-performance liquid chromatography–photodiode detector–quadrupole/time-of-flight–mass spectrometry (UPLC-PDA-TOF-MS) as previously described (Martínez-Pérez et al., 2017). As shown in **Figure 5**, paALKBH5-like protein demethylated m6A on the ssRNA-oligonucleotide substrate by up to 60%, a percentage comparable to that observed for the positive control (ALKBH9B, 75%). Therefore, paALKBH5-like is a protein with ssRNA m6A demethylase activity “*in vitro*.” Previously it was found that change of the histidine at position 204 by alanine (H204A) in the characteristic iron ligand domain of the AlkB family (H204xD206...H266; Lu et al., 2014) of human ALKBH5 completely abolished the demethylation activity of the protein, confirming that the iron center was implicated in this activity (Zheng et al., 2013). The equivalent H204 is located at position 448 (His448, Asp450, and His496) in paALKBH5-like. To check if the putative iron center in paALKBH5-like was required for its demethylation activity, a recombinant version of the protein having the mutation H448A was generated and biochemically tested. We found that the m6A levels were similar to those of the negative control indicating that the change of H448 by A drastically affects the demethylase capacity of paALKBH5-like (**Figure 5**).

Identification of a Putative RNA Binding Site

Recently, a biochemical analysis to delimit functional domains in *Arabidopsis* ALKBH9B showed two intrinsically disordered regions (IDRs) located at N- and C-terminal parts (Alvarado-Marchena et al., 2021). IDRs are unstructured regions implicated in protein–protein and RNA–protein interactions, which among others, present RGG boxes or R/K basic domains (repeats of arginine/glycine domains or arginine- or lysine-rich regions, respectively) [reviewed in Corley et al. (2020) and Ottoz and Berchowitz (2020)]. In ALKBH9B, an RNA binding domain containing an RGxxxRGG overlaps with the C-terminal IDR (Alvarado-Marchena et al., 2021). Both, C-terminal IDR and the RNA binding domain are absent in the TE-encoded protein (ORF2). However, we could detect an IDR in the N-terminal part of ORF2 and we noticed that it contains a basic domain consisting of a path rich with R and K residues (**Supplementary Figure 2**). Arginine residues have been found to interact with the phosphodiester backbone and in this sense RNA binding domain consisting of rich K and R patches have been identified in several RNA binding proteins (Ottoz and Berchowitz, 2020). Thus, is it tempting to speculate that this internal K/R rich domain would confer RNA binding capabilities to the Gypsy_Pa_2799 protein. To check this possibility we carried out Electrophoretic Mobility Shift Assays (EMSA) by incubating a constant amount (5 ng) of an RNA transcript, corresponding to the 5' untranslated region (UTR) of AMV RNA3, with increasing concentrations of the paALKBH5-like (Gypsy_Pa_2799) protein (**Figure 6A**). The decrease in the chemiluminescent signal intensity corresponding to free RNA was evident at quantities exceeding 125 ng of the protein. The apparent constant dissociation (Kd) of the RNA–paALKBH5-like interaction from the linear regression of the mean values from at least three technical replicates (Marcos et al., 1999) was estimated to be 0.21 μ M (**Figure 6B**). To determine if the mutant lacking demethylase activity retained its RNA-binding capacity, Northwestern experiments were performed with both proteins. As can be seen in **Figure 6C**, the His:paALKBH5-like H448A mutant in which its demethylase activity was abolished maintained its RNA binding capacity, indicating that this capacity is necessary but not sufficient for its function.

DISCUSSION

In this study, we screened the content of conserved protein domains in LTR retrotransposon consensus sequences generated from a selection of plant genome assemblies. Admittedly, the set of domains found is biased by the approach used to build consensus sequences, which is empirically not adapted to the analysis of low-copy number TE families and which efficiency decreases with low contiguity genome assemblies. This analysis is also biased by the selection of the species investigated here. Nevertheless, we generated thousands of consensus sequences representative of Copia-type and Gypsy-type LTR retrotransposons and the analysis of their protein domain content allowed to successfully retrieve several domains that were previously reported in plant Gypsy-type elements such

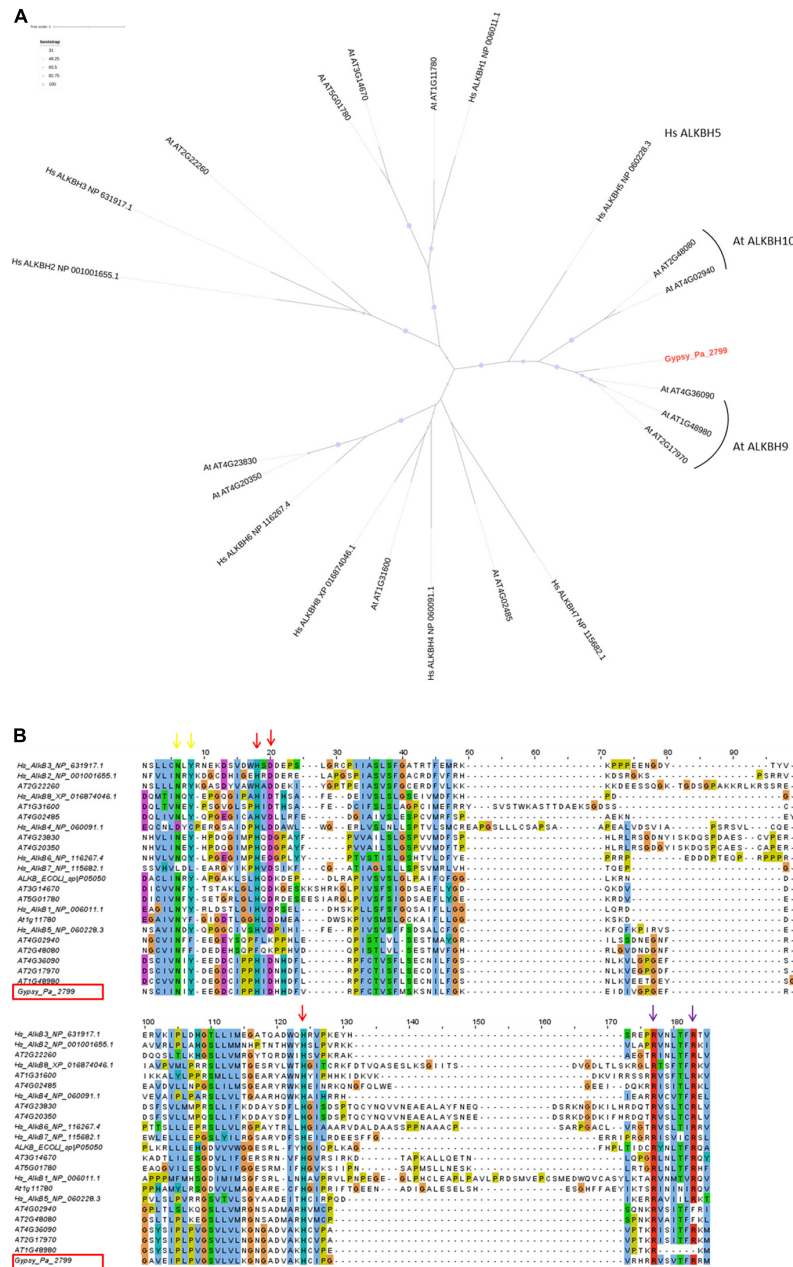
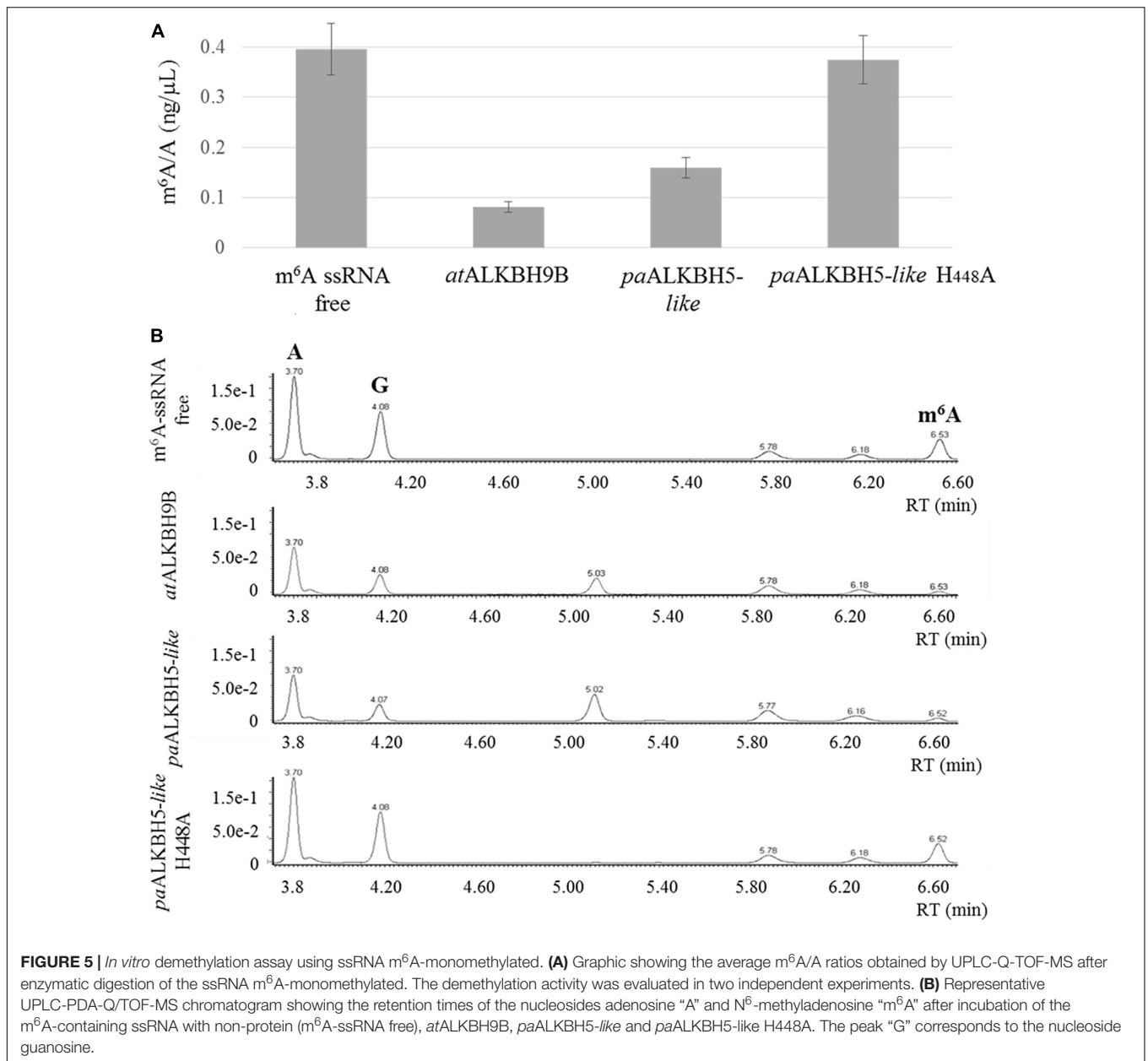


FIGURE 4 | Characterization of the Gypsy-encoded 2-ODD domain. **(A)** Phylogenetic tree inferred from the alignment of Gypsy_Pa_2799 (labeled in red) 2-ODD domain with AlkB homologs from *E. coli*, *A. thaliana*, and *H. sapiens*. The bootstrap values are indicated as circles of size proportional with intervals of values described in the top left panel. **(B)** Conservation of co-factor binding residues in the 2-ODD domain across the alignment used in **(A)**. The alignment view is truncated to positions corresponding to residues 188–285 in human ALKBH5. The columns corresponding to the metal ion binding residues (ALKBH5 His-204, Asp-206, and His-266) are indicated by red arrows. The columns corresponding to the 2-OG binding residues (ALKBH5 Arg-277, and Arg-283) are indicated by purple arrows and the residues involved in 2-OG stabilization (ALKBH5 Asn-193, Tyr-195) are indicated by yellow arrows. The Gypsy_Pa_2799 sequence name is highlighted by a red box.

as CHROMO, PMD, and AIR1. In line with the literature, no domain gain was observed in plant Copia-type elements by contrast to plant Gypsy elements which are more prone to domain acquisition and/or retention.

We noticed that the most frequent protein domains described have functional categories relevant to interactions

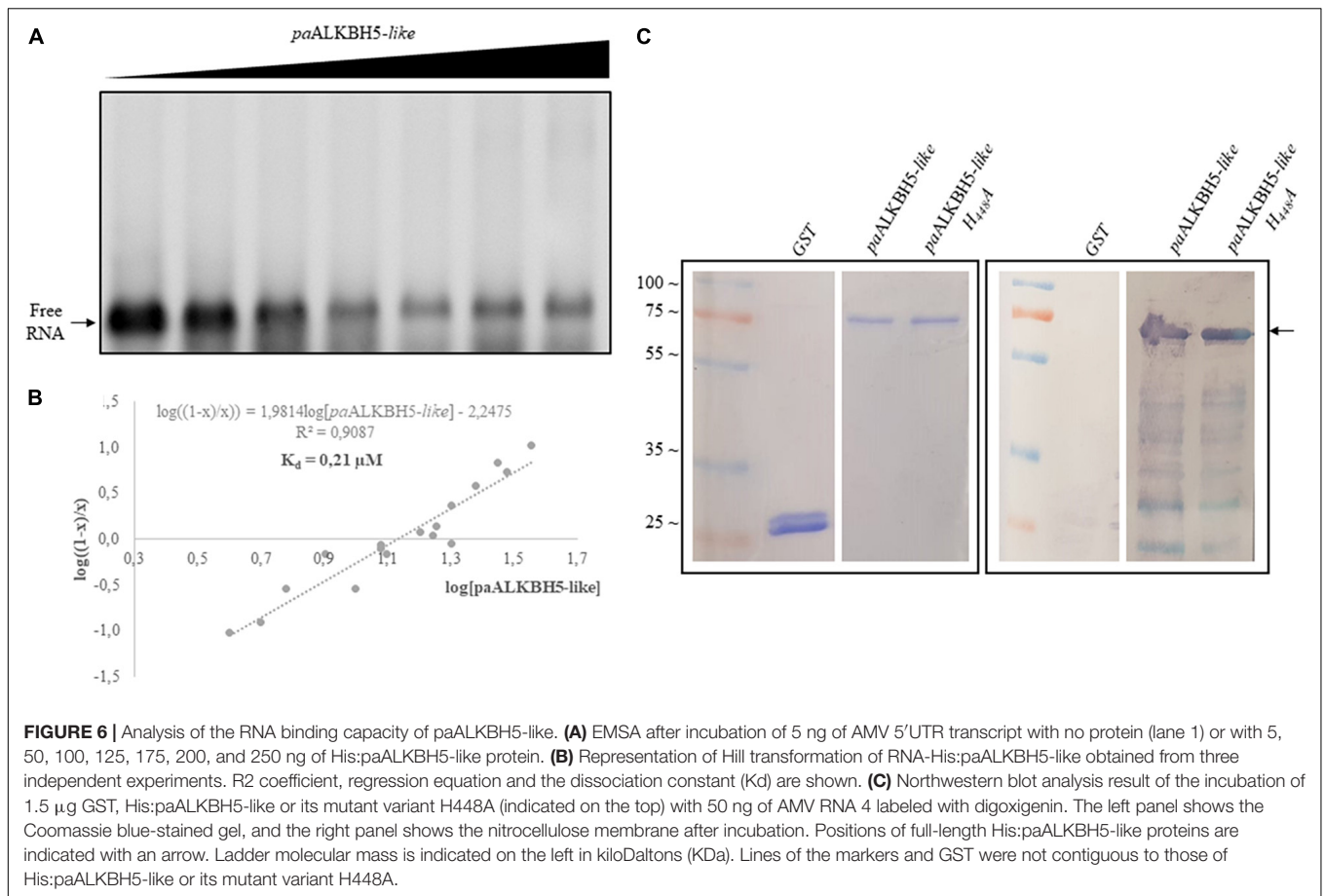
with RNA and DNA including AlkB, G-patch, KOW, ArgoL2, B3, and potentially AIR1 domain, which overlaps UMSBP DNA binding domain. Given that TEs have compact genomes with limited coding capacity, the retention of these domains through evolution and natural selection suggests that they bring important functions to the cognate TE families that result in



a positive impact on TE fitness (with an acceptable impact on host fitness). Consequently, such a functional bias suggests that these fixed domains may have been selected out of a cryptic flow of domain acquisition events, the majority of which being rejected throughout evolution. Remarkably, several of these domains have been acquired independently in different mobile genetic elements. For instance, PMD domain is also found in MULE DNA transposons (Babu et al., 2006). The G-patch has also been detected in type D retroviruses (Aravind and Koonin, 1999). Furthermore, AlkB domain was reported in several families of RNA viruses, including Alphaflexiviridae, Betaflexiviridae, and Closteroviridae (Aravind and Koonin, 2001; Moore and Meng, 2019). The virus-encoded proteins display RNA demethylation activity *in vitro* and it has been postulated

that its RNA repair function could help to maintain genome integrity (van den Born et al., 2008).

Here, we described that a specific family of Phalaenopsis Gypsy-type elements has acquired an additional ORF and retained it over several rounds of replication as attested by the diversity of consensus sequences reconstructed for this family (Figure 2A) which probably reflects speciation into different subfamilies after the ORF was captured. This ORF contains a 2-ODD domain, conserves the cofactor-binding residues necessary for its activity (Figure 4B), is most closely related to known RNA demethylases from human and *Arabidopsis* (Figure 4A) and, most importantly, the purified protein shows RNA demethylase activity *in vitro* (Figure 5) and capability to bind RNA (Figure 6). Interestingly, the apparent constant dissociation (Kd) calculated



for the paALKBH5-like protein (0.21 μ M, **Figure 6**) was similar to that observed for atALKBH9B (0.30 μ M) (Alvarado-Marchena et al., 2021) reinforcing the notion that both proteins are functionally equivalent.

Following the “no superflue” postulate regarding the genomes of functional and autonomous TE sequences, one can suppose that its demethylase activity could bring selective advantage to the TE family itself, e.g., by increasing its fitness. After transcription, the LTR retrotransposon mRNA is exported to the cytoplasm where it is translated (Sabot and Schulman, 2006). We can speculate that, upon translation, the RNA demethylase protein (ORF2), which is expressed independently of the Gag-Pol polyprotein (ORF1), could bind and be enzymatically active against the cognate TE mRNA or any cellular RNA. To bring any advantage to the TE, this could suppose either that the TE mRNA is subject to m6A methylation by the host and that it is deleterious to the TE replication cycle or/and that the demethylation of other cellular RNA by TE-encoded AlkB protein impacts the stability and translation of genes involved in different layers of TE control, as recently proposed for some AlkB-encoding plant RNA viruses (Yue et al., 2022).

m6A methylated mRNAs recruit specific reader proteins and, in mammals, it has been found that a polymethylated mRNA recruits several reader molecules causing the juxtaposition

of their intrinsically disordered regions (IDRs) that activates the phenomenon of liquid-liquid phase separation. Thus, the polymethylated mRNA-eraser protein complexes accumulate into different phase-separated membraneless compartments such as stress granules or P- bodies that affect either the stability or translation of these mRNAs (Ries et al., 2019). Plant m6A readers also contain IDRs and some of them have been found to form dense assemblages *in vitro* and accumulate in granules upon stress (Arribas-Hernandez et al., 2018). Thus, the abundance of m6A sites in TE-derived mRNAs might influence their expression and retrotransposition activities.

Most recently, it has been reported in human cells that RNA transcripts of young LINE-1 retrotransposons present elevated levels of m6A modification that acts promoting their expression and retrotransposition (Xiong et al., 2021). In *Arabidopsis*, most TE transcripts exhibit a relatively high extent of m6A modification compared to gene transcript level (Wan et al., 2015). However, this has been observed under standard conditions in which mostly fragmented forms of TE transcripts can be detected. It would be interesting to measure the level of m6A methylation in TE mRNA in the context of their replicative cycles and to address its impact on the regulation of the transposition rates. Remarkably, a recent report deposited in bioRxiv shows that the RNA of the heat-activated retroelement ONSEN contains m6A,

which suppresses the transposon mobility via its sequestration in stress granules in an m6A-dependent manner. Furthermore, the RNA demethylase AtALKBH9B directly demethylates ONSEN m6A-modified RNA located in stress granules allowing its mobilization. In fact, it has been proposed that in plants, some RNA demethylases would have diversified to preferentially target genetic elements such as retrotransposons (Fan et al., 2022).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Figshare: https://figshare.com/projects/Orchid_Ogre_with_RNA_demethylase_-_Supplementary_dataset/137215.

AUTHOR CONTRIBUTIONS

LA-M and MM-P performed the experiments. VP supervised the experiments. FM performed the bioinformatic analyses and conceived the study. FM, FA, MM-P, and VP interpreted the

results and wrote the manuscript. All authors analyzed and discussed the results.

FUNDING

This research was funded by grant PID2020-115571RB-I00 to FA and VP from the Spanish MCIN/AEI/10.13039/501100011033 granting agency and by financial support from INRAE Plant Biology and Breeding Division to FM. LA-M was recipient of a Predoctoral contract from the Ministerio de Ciencia, Innovación, Tecnología y Telecomunicaciones from Costa Rica (MI-CITT-PINN-CON-624-2019). UPLC-PDAQ/TOF-MS analyses were performed by the Metabolic Analysis Department of the IBMCP.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://figshare.com/projects/Orchid_Ogre_with_RNA_demethylase_-_Supplementary_dataset/137215

REFERENCES

- Alvarado-Marchena, L., Marquez-Molins, J., Martinez-Perez, M., Aparicio, F., and Pallas, V. (2021). Mapping of functional subdomains in the atALKBH9B m(6)A-demethylase required for its binding to the viral RNA and to the Coat Protein of Alfalfa Mosaic Virus. *Front. Plant Sci.* 12:701683. doi: 10.3389/fpls.2021.701683
- Anderson, S. J., Kramer, M. C., Gosai, S. J., Yu, X., Vandivier, L. E., Nelson, A. D. L., et al. (2018). N(6)-Methyladenosine inhibits local ribonucleolytic cleavage to stabilize mRNAs in *Arabidopsis*. *Cell Rep.* 25, 1146–1157 e3. doi: 10.1016/j.celrep.2018.10.020
- Aravind, L., and Koonin, E. V. (1999). G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem. Sci.* 24, 342–344.
- Aravind, L., and Koonin, E. V. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* 2:RESEARCH0007. doi: 10.1186/gb-2001-2-3-research0007
- Arkhipova, I. R., and Yushenova, I. A. (2019). Giant transposons in eukaryotes: is bigger better? *Genome Biol. Evol.* 11, 906–918. doi: 10.1093/gbe/evz041
- Arribas-Hernandez, L., Bressendorff, S., Hansen, M. H., Poulsen, C., Erdmann, S., and Brodersen, P. (2018). An m(6)A-YTH module controls developmental timing and Morphogenesis in *Arabidopsis*. *Plant Cell* 30, 952–967. doi: 10.1105/tpc.17.00833
- Arribas-Hernandez, L., and Brodersen, P. (2020). Occurrence and functions of m(6)A and Other covalent modifications in plant mRNA. *Plant Physiol.* 182, 79–96. doi: 10.1104/pp.19.01156
- Babu, M. M., Iyer, L. M., Balaji, S., and Aravind, L. (2006). The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* 34, 6505–6520. doi: 10.1093/nar/gkl888
- Bratlie, M. S., and Drablos, F. (2005). Bioinformatic mapping of AlkB homology domains in viruses. *BMC Genomics* 6:1. doi: 10.1186/1471-2164-6-1
- Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741.
- Chelmicki, T., Roger, E., Teissandier, A., Dura, M., Bonneville, L., Ruclis, S., et al. (2021). m(6)A RNA methylation regulates the fate of endogenous retroviruses. *Nature* 591, 312–316. doi: 10.1038/s41586-020-03135-1
- Corley, M., Burns, M. C., and Yeo, G. W. (2020). How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* 78, 9–29.
- Duan, H. C., Wei, L. H., Zhang, C., Wang, Y., Chen, L., Lu, Z., et al. (2017). ALKBH10B Is an RNA N(6)-methyladenosine demethylase affecting *Arabidopsis* floral transition. *Plant Cell* 29, 2995–3011. doi: 10.1105/tpc.16.00912
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18. doi: 10.1186/1471-2105-9-18
- Faehnle, C. R., Elkayam, E., Haase, A. D., Hannon, G. J., and Joshua-Tor, L. (2013). The making of a slicer: activation of human Argonaute-1. *Cell Rep.* 3, 1901–1909. doi: 10.1016/j.celrep.2013.05.033
- Fan, W., Wang, L., Lei, Z., Chu, J., and Cho, J. (2022). Suppression of transposon mobilization by m6A-mediated RNA sequestration in stress granules. *bioRxiv* [Preprint] doi: 10.1101/2022.03.22.485398
- Farrow, S. C., and Facchini, P. J. (2014). Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front. Plant Sci.* 5:524. doi: 10.3389/fpls.2014.00524
- Feng, C., Liu, Y., Wang, G., Deng, Z., Zhang, Q., Wu, W., et al. (2014). Crystal structures of the human RNA demethylase Alkbh5 reveal basis for substrate recognition. *J. Biol. Chem.* 289, 11571–11583. doi: 10.1074/jbc.M113.546168
- Gao, X., Hou, Y., Ebina, H., Levin, H. L., and Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res.* 18, 359–369. doi: 10.1101/gr.7146408
- Gonzales-van Horn, S. R., and Sarnow, P. (2017). Making the mark: the role of adenosine modifications in the life cycle of RNA viruses. *Cell Host Microbe* 21, 661–669. doi: 10.1016/j.chom.2017.05.008
- Gorinsek, B., Gubensek, F., and Kordis, D. (2004). Evolutionary genomics of chromoviruses in eukaryotes. *Mol. Biol. Evol.* 21, 781–798.
- He, Y., Li, Y., Yao, Y., Zhang, H., Wang, Y., Gao, J., et al. (2021). Overexpression of watermelon m(6)A methyltransferase CLMTB enhances drought tolerance in tobacco by mitigating oxidative stress and photosynthesis inhibition and modulating stress-responsive gene expression. *Plant Physiol. Biochem.* 168, 340–352. doi: 10.1016/j.plaphy.2021.10.007
- Herr, C. Q., and Hausinger, R. P. (2018). Amazing diversity in biochemical roles of Fe(II)/2-oxoglutarate oxygenases. *Trends Biochem. Sci.* 43, 517–532. doi: 10.1016/j.tibs.2018.04.002
- Hu, J., Cai, J., Park, S. J., Lee, K., Li, Y., Chen, Y., et al. (2021a). N(6)-Methyladenosine mRNA methylation is important for salt stress tolerance in *Arabidopsis*. *Plant J.* 106, 1759–1775. doi: 10.1111/tjpi.15270

- Hu, J., Cai, J., Umme, A., Chen, Y., Xu, T., and Kang, H. (2021b). Unique features of mRNA m6A methylomes during expansion of tomato (*Solanum lycopersicum*) fruits. *Plant Physiol.* 3:kiab509. doi: 10.1093/plphys/kiab509
- Hu, J., Manduzio, S., and Kang, H. (2019). Epitranscriptomic RNA methylation in plant development and abiotic stress responses. *Front. Plant Sci.* 10:500. doi: 10.3389/fpls.2019.00500
- Ikeda, Y., Pelissier, T., Bourguet, P., Becker, C., Pouch-Pelissier, M. N., Pogorelnik, R., et al. (2017). *Arabidopsis* proteins with a transposon-related domain act in gene silencing. *Nat. Commun.* 8:15122. doi: 10.1038/ncomms15122
- Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kawai, Y., Ono, E., and Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* 78, 328–343.
- Kazanian, H. H. Jr. (2004). Mobile elements: drivers of genome evolution. *Science* 303, 1626–1632.
- Kuroha, T., Tokunaga, H., Kojima, M., Ueda, N., Ishida, T., Nagawa, S., et al. (2009). Functional analyses of LONELY GUY cytokinin-activating enzymes reveal the importance of the direct activation pathway in *Arabidopsis*. *Plant Cell* 21, 3152–3169. doi: 10.1105/tpc.109.068676
- Kurowski, M. A., Bhagwat, A. S., Papaj, G., and Bujnicki, J. M. (2003). Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* 4:48. doi: 10.1186/1471-2164-4-48
- Kyrpides, N. C., Woese, C. R., and Ouzounis, C. A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* 21, 425–426. doi: 10.1016/s0968-0004(96)30036-4
- Laten, H. M., Majumdar, A., and Gaucher, E. A. (1998). SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6897–6902. doi: 10.1073/pnas.95.12.6897
- Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 518, 560–564. doi: 10.1038/nature14234
- Llorens, C., Futami, R., Covelli, L., Dominguez-Escriba, L., Viu, J. M., Tamarit, D., et al. (2011). The gypsy database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* 39, D70–D74.
- Llorens, C., Munoz-Pomer, A., Bernad, L., Botella, H., and Moya, A. (2009). Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol. Direct* 4:41. doi: 10.1186/1745-6150-4-41
- Lu, L., Zhu, C., Xia, B., and Yi, C. (2014). Oxidative demethylation of DNA and RNA mediated by non-heme iron-dependent dioxygenases. *Chem. Asian J.* 9, 2018–2029. doi: 10.1002/asia.201402148
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2010). CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229.
- Marcos, J. F., Vilar, M., Perez-Paya, E., and Pallas, V. (1999). In vivo detection, RNA-binding properties and characterization of the RNA-binding domain of the p7 putative movement protein from carnation mottle carmovirus (CarMV). *Virology* 255, 354–365. doi: 10.1006/viro.1998.9596
- Marin, I., and Llorens, C. (2000). Ty3/Gypsy retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. *Mol. Biol. Evol.* 17, 1040–1049. doi: 10.1093/oxfordjournals.molbev.a026385
- Martínez-Pérez, M., Aparicio, F., Lopez-Gresa, M. P., Belles, J. M., Sanchez-Navarro, J. A., and Pallas, V. (2017). *Arabidopsis* m(6)A demethylase activity modulates viral infection of a plant virus and the m(6)A abundance in its genomic RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10755–10760. doi: 10.1073/pnas.1703139114
- Martínez-Pérez, M., Gomez-Mena, C., Alvarado-Marchena, L., Nadi, R., Micol, J. L., Pallas, V., et al. (2021). The m(6)A RNA demethylase ALKBH9B plays a critical role for vascular movement of alfalfa mosaic virus in *Arabidopsis*. *Front. Microbiol.* 12:745576. doi: 10.3389/fmicb.2021.745576
- Matzke, M. A., Kanno, T., and Matzke, A. J. (2015). RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annu. Rev. Plant Biol.* 66, 243–267.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534.
- Moore, C., and Meng, B. (2019). Prediction of the molecular boundary and functionality of novel viral AlkB domains using homology modelling and principal component analysis. *J. Gen. Virol.* 100, 691–703. doi: 10.1099/jgv.0.001237
- Nayar, S. (2020). Exploring the role of a cytokinin-activating enzyme LONELY GUY in unicellular microalga *Chlorella variabilis*. *Front. Plant Sci.* 11:611871. doi: 10.3389/fpls.2020.611871
- Neumann, P., Novák, P., Hošťáková, N., and Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA* 10:1. doi: 10.1186/s13100-018-0144-1
- Neumann, P., Pozarkova, D., and Macas, J. (2003). Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. *Plant Mol. Biol.* 53, 399–410. doi: 10.1023/b:plan.0000006945.77043.ce
- Nicolau, M., Picault, N., Descombin, J., Jami-Alahmadi, Y., Feng, S., Bucher, E., et al. (2020). The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*. *PLoS Genet.* 16:e1008324. doi: 10.1371/journal.pgen.1008324
- Ottoz, D. S. M., and Berchowitz, L. E. (2020). The role of disorder in RNA binding affinity and specificity. *Open Biol.* 10:200328.
- Pallas, V., Sanchez-Navarro, J. A., and Diez, J. (1999). In vitro evidence for RNA binding properties of the coat protein of prunus necrotic ringspot ilarvirus and their comparison to related and unrelated viruses. *Arch. Virol.* 144, 797–803. doi: 10.1007/s007050050545
- Ries, R. J., Zaccara, S., Klein, P., Olarerier-George, A., Namkoong, S., Pickering, B. F., et al. (2019). m(6)A enhances the phase separation potential of mRNA. *Nature* 571, 424–428.
- Ruzicka, K., Zhang, M., Campilho, A., Bodi, Z., Kashif, M., Saleh, M., et al. (2017). Identification of factors required for m(6) A mRNA methylation in *Arabidopsis* reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytol.* 215, 157–172. doi: 10.1111/nph.14586
- Sabot, F., and Schulman, A. H. (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity (Edinb)* 97, 381–388. doi: 10.1038/sj.hdy.6800903
- Saint-Leandre, B., Nguyen, S. C., and Levine, M. T. (2019). Diversification and collapse of a telomere elongation mechanism. *Genome Res.* 29, 920–931. doi: 10.1101/gr.245001.118
- Shen, L., Liang, Z., Gu, X., Chen, Y., Teo, Z. W., Hou, X., et al. (2016). N(6)-methyladenosine RNA modification regulates shoot stem cell fate in *Arabidopsis*. *Dev. Cell* 38, 186–200. doi: 10.1016/j.devcel.2016.06.008
- Steinbauerova, V., Neumann, P., Novak, P., and Macas, J. (2011). A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. *Genetica* 139, 1543–1555. doi: 10.1007/s10709-012-9654-9
- Steinberger, M., and Soding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. doi: 10.1038/nbt.3988
- Ustyantsev, K., Novikova, O., Blinov, A., and Smyshlyayev, G. (2015). Convergent evolution of ribonuclease h in LTR retrotransposons and retroviruses. *Mol. Biol. Evol.* 32, 1197–1207. doi: 10.1093/molbev/msv008
- van den Born, E., Omelchenko, M. V., Bekkelund, A., Leihne, V., Koonin, E. V., Dolja, V. V., et al. (2008). Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic Acids Res.* 36, 5451–5461.
- Vicient, C. M., and Casacuberta, J. M. (2020). Additional ORFs in Plant LTR-retrotransposons. *Front. Plant Sci.* 11:555. doi: 10.3389/fpls.2020.00555
- Wan, Y., Tang, K., Zhang, D., Xie, S., Zhu, X., Wang, Z., et al. (2015). Transcriptome-wide high-throughput deep m(6)A-seq reveals unique differential m(6)A methylation patterns between three organs in *Arabidopsis thaliana*. *Genome Biol.* 16:272. doi: 10.1186/s13059-015-0839-2
- Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505, 117–120.
- Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161, 1388–1399.

- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhou, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982.
- Williams, G. D., Gokhale, N. S., and Horner, S. M. (2019). Regulation of viral infection by the RNA modification N6-methyladenosine. *Annu. Rev. Virol.* 6, 235–253.
- Wright, D. A., and Voytas, D. F. (1998). Potential retroviruses in plants: tat1 is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. *Genetics* 149, 703–715. doi: 10.1093/genetics/149.2.703
- Xiong, F., Wang, R., Lee, J. H., Li, S., Chen, S. F., Liao, Z., et al. (2021). RNA m(6A) modification orchestrates a LINE-1-host interaction that facilitates retrotransposition and contributes to long gene vulnerability. *Cell Res.* 31, 861–885. doi: 10.1038/s41422-021-00515-8
- Xu, C., Liu, K., Tempel, W., Demetriades, M., Aik, W., Schofield, C. J., et al. (2014). Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single-stranded N6-methyladenosine RNA demethylation. *J. Biol. Chem.* 289, 17299–17311. doi: 10.1074/jbc.M114.550350
- Yu, Q., Liu, S., Yu, L., Xiao, Y., Zhang, S., Wang, X., et al. (2021). RNA demethylation increases the yield and biomass of rice and potato plants in field trials. *Nat. Biotechnol.* 39, 1581–1588. doi: 10.1038/s41587-021-00982-9
- Yue, H., Nie, X., Yan, Z., and Weining, S. (2019). N6-methyladenosine regulatory machinery in plants: composition, function and evolution. *Plant Biotechnol. J.* 17, 1194–1208. doi: 10.1111/pbi.13149
- Yue, J., Wei, Y., and Zhao, M. (2022). The reversible methylation of m6A is involved in plant virus infection. *Biology (Basel)* 11:271.
- Zaccara, S., and Jaffrey, S. R. (2020). A unified model for the function of YTHDF proteins in regulating m(6A)-modified mRNA. *Cell* 181:e1518. doi: 10.1016/j.cell.2020.05.012
- Zheng, H., Sun, X., Li, J., Song, Y., Song, J., Wang, F., et al. (2021). Analysis of N(6)-methyladenosine reveals a new important mechanism regulating the salt tolerance of sweet sorghum. *Plant Sci.* 304:110801. doi: 10.1016/j.plantsci.2020.110801
- Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29. doi: 10.1016/j.molcel.2012.10.015
- Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., et al. (2008). MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* 20, 1278–1288. doi: 10.1105/tpc.108.058883

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Alvarado-Marchena, Martínez-Pérez, Aparicio, Pallas and Maumus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.